# Fast factorization-based inference for Bayesian harmonic models

Emmanuel Vincent, Mark Plumbley

# FAST FACTORIZATION-BASED INFERENCE FOR BAYESIAN HARMONIC MODELS

*Emmanuel Vincent and Mark D. Plumbley*

Centre for Digital Music
Department of Electronic Engineering, Queen Mary, University of London
Mile End Road, London E1 4NS, United Kingdom

## ABSTRACT

Harmonic sinusoidal models are a fundamental tool for audio signal analysis. Bayesian harmonic models guarantee a good resynthesis quality and allow joint use of learnt parameter priors and auditory motivated distortion measures. However inference algorithms based on Monte Carlo sampling are rather slow for realistic data. In this paper, we investigate fast inference algorithms based on approximate factorization of the joint posterior into a product of independent distributions on small subsets of parameters. We discuss the conditions under which these approximations hold true and evaluate their performance experimentally. We suggest how they could be used together with Monte Carlo algorithms for a faster sampling-based inference.

## 1. INTRODUCTION

Music and speech involve different types of sounds, including periodic, transient and noisy sounds. Short-term stationary periodic sounds composed of sinusoidal partials at harmonic frequencies are particularly important perceptually, since they represent most of the energy of musical notes and vowels. Harmonicity means that at each instant the frequencies of the partials are multiples of a single frequency called the fundamental frequency. Estimating the periodic sounds underlying a given signal, *i.e.* estimating their fundamental frequencies and the amplitudes and phases of their partials, is required or useful for many applications, such as audio indexing, browsing by content, source separation, low bitrate compression, musical score transcription and interactive content manipulation. This problem is particularly difficult for polyphonic signals, *i.e.* signals containing several concurrent periodic sounds, since different periodic sounds may exhibit partials overlapping at the same frequencies.

Existing methods for polyphonic fundamental frequency estimation are often based on one of two approaches: either validation of fundamental frequency candidates given by the peaks of a short-term auto-correlation function [1, 2] or inference of the hidden states of a probabilistic model of the signal short-term power spectrum using detailed prior information [3, 4]. These approaches have achieved a limited performance on complex polyphonic excerpts so far [2]. Moreover neither approach estimates the parameters of the partials, which are needed for some applications.

A promising way to address these issues is to rely on a probabilistic model of the signal waveform involving fundamental frequency, amplitude and phase parameters. A family of such models has been proposed in the literature for music signals, along with Markov Chain Monte Carlo (MCMC) methods to infer their parameters [5, 6, 7]. These methods converge to the right solution asymptotically, but tend to be rather slow on realistic examples [8]. Thus the parameter priors chosen in these models are partly motivated by computational issues. For instance, the prior over the number of partials per note favors a small number of partials independently of the fundamental frequency in [5, 6], while the amplitudes of the partials are modeled by conjugate priors [8] such as a uniform prior in [5] and zero-mean Gaussian priors with various covariance matrices in [6, 7]. These priors do not penalize partials with zero amplitude, which can lead to erroneous fundamental frequency estimates or bad quality separated note signals. To help solving these limitations, we recently proposed a harmonic model [9] including probabilistic priors motivated by observation of empirical parameter distributions and used the diagonal Laplace method [10] for fast parameter inference.

In this paper, we develop improved fast inference methods for Bayesian harmonic models, based on factorization of the joint posterior into a product of independent distributions on subsets of parameters. These methods are illustrated in the particular case of the proposed model, but could be applied to some other types of harmonic models.

The structure of the rest of the paper is as follows. In section 2, we briefly introduce our harmonic model and the associated parameter priors. Then, we describe the proposed inference methods in section 3 and discuss the conditions under which they give a precise result. In section 4, we evaluate their performance for musical score transcription on short time frames. We conclude in section 5 and suggest a way of combining these methods with MCMC.

This work is funded by the EPSRC grant GR/S75802/01.

## 2. MODEL DEFINITION

The model proposed in [9] represents a music signal as a collection of notes, each composed of harmonic sinusoidal partials. For simplicity, we assume in the following that parameters on different time frames are independent. On each time frame, the model exhibits the four-layer Bayesian network structure shown in figure 1. Each layer models the observed signal frame $x(t)$ at a different abstraction level.

### 2.1. Structure of the proposed model

The bottom layer represents the underlying musical score. In western music, the normalized fundamental frequency $f_{pn}$ of each note may vary but remains close to a discrete pitch of the form

$$\mu_p^f = \frac{440}{F_s} 2^{\frac{p-69}{12}} \tag{1}$$

where $F_s$ is the sampling frequency and $p$ an integer value on the MIDI semitone scale. Each discrete pitch $p$ is associated with a binary state $S_p$ determining whether a note with that discrete pitch is active or not. The signal $s_p(t)$ corresponding to each active note is then defined in the middle layers for $0 \leq t \leq T-1$ by

$$s_p(t) = w(t) \sum_{m=1}^{M_p} a_{pm} \cos(2\pi m f_p t + \phi_{pm}), \tag{2}$$

where $w(t)$ is the framing window and $f_p$, $a_{pm}$ and $\phi_{pm}$ are respectively its normalized fundamental frequency and the amplitude and the phase of its $m$-th partial. The number of partials $M_p$ is constrained to $M_p = \min(1/(2\mu_p^f), M_{\max})$. Finally, the observed signal is modeled in the top layer as

$$x(t) = \sum_{p/S_p=1} s_p(t) + e(t), \tag{3}$$

where $e(t)$ is the residual.

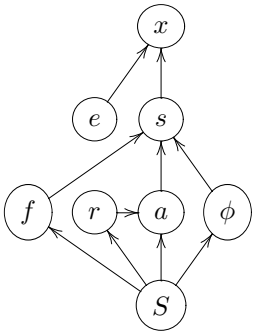
### 2.2. State and parameter priors

In order to penalize transcriptions containing too many active notes, the global state $S = (S_p)_{P \leq p \leq P'}$ is modeled by a product of independent Bernoulli distributions

$$P(S) = \prod_{p/S_p=1} (1 - P_Z) \prod_{p/S_p=0} P_Z, \tag{4}$$

where $P_Z$ is the mean inactivity probability. Given the note states, the parameters of different notes are assumed to be independent. The normalized fundamental frequency of each note is modeled by a log-Gaussian prior enforcing proximity to the underlying discrete pitch

$$P(\log f_p) = \mathcal{N}(\log f_p; \log \mu_p^f, \sigma^f), \tag{5}$$




$x$: observed signal frame

$(s_p)$: note signals
$e$: residual

$(f_p)$: fundamental frequencies
$(r_p)$: global amplitude factors
$(a_{pm})$: amplitudes of the partials
$(\phi_{pm})$: phases of the partials

$(S_p)$: note states

**Fig. 1**. Graphical representation of the proposed model. Subscripts are omitted for legibility.

where $\mathcal{N}(\cdot; \mu, \sigma)$ is the univariate Gaussian density of mean $\mu$ and standard deviation $\sigma$. Similarly, the amplitudes of the partials are represented as

$$P(\log a_{pm} | r_p) = \mathcal{N}(\log a_{pm}; \log(r_p \mu_{pm}^a), \sigma_p^a), \tag{6}$$

where $(\mu_{pm}^a)_{1 \leq m \leq M_p}$ is a fixed (learnt) normalized spectral envelope helping to avoid partials with zero amplitude and $r_p$ a global amplitude factor for this note, modeled by

$$P(\log r_p) = \mathcal{N}(\log r_p; \log \mu_p^r, \sigma_p^r). \tag{7}$$

The phases of the partials are assumed to be independent and uniformly distributed

$$P(\phi_{pm}) = 1/2\pi. \tag{8}$$

### 2.3. Distortion measure

The distortion between the observed signal and the model is measured by the auditory motivated weighted Euclidean norm $D = \sum_{f=0}^{T-1} \gamma_f |E_f|^2$, where $(E_f)_{0 \leq f \leq T-1}$ are the discrete Fourier transform coefficients of $e(t)$ and the constant frequency weights $(\gamma_f)_{0 \leq f \leq T-1}$ are given in [9]. The residual prior is derived by $P(e) \propto \exp(-D/(2(\sigma^e)^2))$, resulting in the weighted Gaussian distribution

$$P(e) = \prod_{f=0}^{T-1} \mathcal{N}(E_f; 0, \sigma^e \gamma_f^{-1/2}). \tag{9}$$

## 3. INFERENCE ALGORITHMS

The aim of musical score transcription is to estimate the Maximum A Posteriori (MAP) state $\hat{S} = \arg\max P(S|x)$. The posterior probability of $S$ equals the integral $P(S|x) = \int P(S, f, r, a, \phi | x)\, df\, dr\, da\, d\phi$, where the joint posterior is expressed by Bayes law $P(S, f, r, a, \phi | x) \propto P(e) P(a | r, S)$ $P(\phi | S) P(r | S) P(f | S) P(S)$. The computation of this integral is known as the Bayesian marginalization problem [8].

Numerical integration is intractable since the number of parameters is typically of the order of one hundred per frame.

Fast inference can be achieved by estimating the MAP parameters $(\hat{f}, \hat{r}, \hat{a}, \hat{\phi}) = \arg\max P(\hat{S}, f, r, a, \phi|x)$ using a standard optimization algorithm[1] and approximating the joint posterior around these values by a simpler distribution which can be integrated analytically. Popular methods include the full Laplace approximation [10], which replaces the posterior by a Gaussian distribution with full covariance matrix, and the diagonal Laplace approximation [10], which factorizes the posterior into a product of parameter-wise univariate Gaussian distributions. Both approximations are better used on log-parameters $\log f_p$, $\log r_p$ and $\log a_{pm}$, which are unbounded [10]. The diagonal Laplace approximation also allows bounded integration over each phase parameter $\phi_{pm}$ in $[-\pi, \pi]$.

The proposed inference methods generalize the diagonal Laplace approximation by factorizing the posterior as a product of Gaussian and non-Gaussian distributions. These factorizations are obtained in several steps.

### 3.1. Conditional posterior factorization over the partials

Let us assume initially that the harmonic partials corresponding to the hypothesized fundamental frequencies have "different enough" frequencies. This is true for a single hypothesized note, but generally not for several notes. Mathematically, this translates into the fact that the windowed complex sinusoidal signals

$$z_{pm}(t) = w(t)e^{2i\pi m f_p t} \tag{10}$$

corresponding to different partials are mutually orthogonal

$$\langle z_{pm}, z_{p'm'} \rangle = 0 \quad \forall (p,m) \neq (p',m') \tag{11}$$

according to the dot product consistent with the distortion measure $D$, defined for two signals $z(t)$ and $z'(t)$ by

$$\langle z, z' \rangle = \sum_{f=0}^{T-1} \gamma_f Z_f \bar{Z}'_f, \tag{12}$$

where $Z_f$ and $Z'_f$ are the discrete Fourier transform coefficients of $z(t)$ and $z'(t)$ and $\bar{Z}'_f$ is the complex conjugate of $Z'_f$. This orthogonality property does not depend on the value of the frequency weights. When the frequencies of the partials are not too close to Nyquist, the negative frequency sinusoidal signals $\bar{z}_{pm}(t) = w(t)e^{-2i\pi m f_p t}$ are also orthogonal to their positive counterparts: $\langle z_{pm}, \bar{z}_{p'm'} \rangle = 0$ for all $(p,m)$ and $(p',m')$. The observed signal $x(t)$ can then be decomposed into a sum of sinusoidal signals at the frequencies of the hypothesized partials by orthogonal projection onto the two-dimensional subspaces spanned by $(z_{pm}, \bar{z}_{pm})$

$$x(t) = \frac{1}{2} \sum_{p,m} \tilde{a}_{pm}(e^{i\tilde{\phi}_{pm}} z_{pm}(t) + e^{-i\tilde{\phi}_{pm}} \bar{z}_{pm}(t)) + \tilde{e}(t). \tag{13}$$

The projection coefficients, given by

$$\tilde{a}_{pm} e^{i\tilde{\phi}_{pm}} = 2\langle x, z_{pm} \rangle / \|z_{pm}\|^2 \tag{14}$$

with $\|z\|^2 = \langle z, z \rangle$, represent the amplitude and phase values of each partial resulting in the minimal distortion. Given hypothesized values $a_{pm}$ and $\phi_{pm}$, the residual $e(t)$ can be decomposed as a sum of mutually orthogonal terms

$$\begin{aligned} e(t) = \frac{1}{2} \sum_{p,m} &\left( \tilde{a}_{pm} e^{i\tilde{\phi}_{pm}} - a_{pm} e^{i\phi_{pm}} \right) z_{pm}(t) \\ &+ \left( \tilde{a}_{pm} e^{-i\tilde{\phi}_{pm}} - a_{pm} e^{-i\phi_{pm}} \right) \bar{z}_{pm}(t) + \tilde{e}(t). \end{aligned} \tag{15}$$

The resulting distortion $D = \|e\|^2$ then equals by analytical computation

$$D = \sum_{p,m} D_{pm} + D_0 \tag{16}$$

where $D_0 = \|\tilde{e}\|^2$ and

$$\begin{aligned} D_{pm} = \frac{1}{2} \|z_{pm}\|^2 \Big( &(a_{pm} - \tilde{a}_{pm})^2 \\ &+ 4\tilde{a}_{pm} a_{pm} \sin^2((\phi_{pm} - \tilde{\phi}_{pm})/2) \Big). \end{aligned} \tag{17}$$

This decomposition results in the factorization of the posterior as a product of partial-wise bivariate conditional distributions over amplitude and phase parameters

$$\begin{aligned} P(S, f, r, a, \phi|x) \propto\ & P_0(x, f) P(r|S) P(f|S) P(S) \\ & \times \prod_{p,m} P_{pm}(a_{pm}, \phi_{pm}; x, f_p) P(a_{pm}|r_p) P(\phi_{pm}), \end{aligned} \tag{18}$$

where $P_0(x, f) = (2\pi\sigma^{e\,2})^{-T/2}\, e^{-D_0/(2\sigma^{e\,2})}$ is a constant and $P_{pm}(a_{pm}, \phi_{pm}; x, f_p) = \exp(-D_{pm}/(2\sigma^{e\,2}))$ a bivariate parametric distribution depending on three hyper-parameters only: $\|z_{pm}\|^2$, $\tilde{a}_{pm}$ and $\tilde{\phi}_{pm}$. The top plot of figure 2 illustrates the validity of this factorization.

### 3.2. Validity of the conditional factorization

In the general case where several partials may have close frequencies, the amplitude and phase values $\tilde{a}_{pm}$ and $\tilde{\phi}_{pm}$ minimizing the distortion for each partial given the MAP values of other parameters can be obtained similarly by

$$\tilde{a}_{pm} e^{i\tilde{\phi}_{pm}} = 2\langle \hat{e}, z_{pm} \rangle / \|z_{pm}\|^2 + \hat{a}_{pm} e^{i\hat{\phi}_{pm}}, \tag{19}$$

where $\hat{e}(t)$ is the residual corresponding to the MAP parameters $(\hat{a}, \hat{\phi})$ depending on $f$ and $r$. Consequently, the quantities $D_0$, $D_{pm}$, $P_0(x, f)$ and $P_{pm}(a_{pm}, \phi_{pm}; x, f_p)$ can still

[1] In the following, we use Matlab's lsqnonlin function.

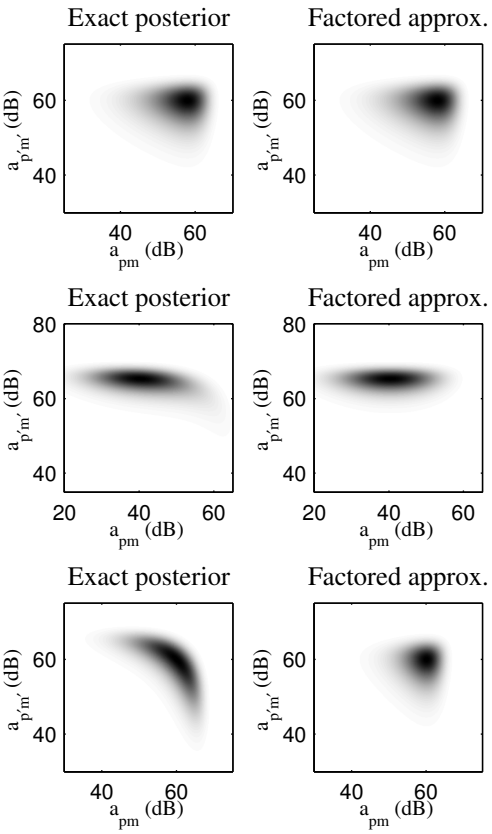



**Fig. 2**. Shape of the joint posterior for a signal containing two partials $(p, m)$ and $(p', m')$ with 60 dB amplitudes as a function of their hypothesized amplitudes. Dark areas denote high probability. Top: partials with different frequencies and mean prior amplitudes of 50 dB and 60 dB. Middle: partials with the same frequency but different mean prior amplitudes of 40 dB and 60 dB. Bottom: partials with the same frequency and same mean prior amplitudes of 60 dB.

be computed. However, the above factorization of the posterior does not always hold true.

This factorization remains valid in the particular case where several partials are located at the same frequency, but only one (masker) has a mean prior amplitude $r_p \mu^a_{pm}$ close to the observed amplitude, while others (masked) have much smaller mean prior amplitudes. Indeed, on the one hand, when the hypothesized amplitude values are far from their prior values, the posterior is small because the factors $P(a_{pm}|r_p)$ are small and the error on $P(e)$ due to factorization does not matter. On the other hand, when the hypothesized amplitudes are close to their prior values, the distortion remains relatively constant as a function of the amplitudes of the masked partials. Thus $P(e)$ depends only on the amplitude of the masker partial and factorizes as above with $P_{pm}(a_{pm}, \phi_{pm}; x, f_p) = 1$ for the masked partials. The bottom plots of figure 2 illustrate this in the case of two partials. The partial-wise factorization appears to be valid when the mean prior amplitude of the masked partial is 20 dB below the amplitude of the masker partial, but not when both amplitudes are equal.

In the latter case, it is still possible to group partials into subsets according to their frequencies such that partials from different subsets are orthogonal, but partials within each subset are not. Similar arguments as above then lead to factorize the posterior as a product of multivariate conditional distributions over subsets of amplitude and phase parameters $a_g = (a_{pm})_{(p,m)\in g}$ and $\phi_g = (\phi_{pm})_{(p,m)\in g}$

$$P(S, f, r, a, \phi|x) \propto P_0(x, f)P(r|S)P(f|S)P(S) \times \prod_g P_g(a_g, \phi_g; x, f)P(a_g|r)P(\phi_g). \quad (20)$$

### 3.3. Full posterior factorization over the partials or subsets of partials

The exact conditional factorizations in equations (18) and (20) can be exploited for numerical integration of the posterior. Indeed integration over amplitude and phase parameters can be achieved by multiplying lower dimension integrals over the parameters of each partial or each subset of partials. Denoting by $N$ the number of grid points for each scalar variable, $P$ the number of hypothesized notes and $M = \sum_p M_p$ their total number of partials, equation (18) requires $N^2$ evaluations of the posterior for each partial and each value of $f$ and $r$, resulting in a complexity of $\mathcal{O}(MN^{2P+2})$. Similarly, equation (20) results at most in a complexity of $\mathcal{O}(\frac{M}{P}N^{4P})$. This is faster than the complexity of $\mathcal{O}(N^{2P+2M})$ associated with straightforward integration, but still intractable.

In order to get faster integration, it is necessary to replace conditional factorization over amplitude and phase parameters by full factorization. An approximate solution is to replace the free parameters in the expression of the conditional distributions by their MAP values. This gives

$$P(S, f, r, a, \phi|x) \approx P_0(x, \hat{f})P(r|S, \hat{f}, \hat{a}, \hat{\phi})P(f|S, \hat{a}, \hat{\phi}) \times P(S) \prod_{p,m} P_{pm}(a_{pm}, \phi_{pm}; x, \hat{f}_p)P(a_{pm}|\hat{r}_p)P(\phi_{pm}) \quad (21)$$

when the partials have "different enough" frequencies and

$$P(S, f, r, a, \phi|x) \approx P_0(x, \hat{f})P(r|S, \hat{f}, \hat{a}, \hat{\phi})P(f|S, \hat{a}, \hat{\phi}) \times P(S) \prod_g P_g(a_g, \phi_g; x, \hat{f})P(a_g|\hat{r})P(\phi_g) \quad (22)$$

in the general case. These equations allow approximate numerical integration of the posterior with a complexity of $\mathcal{O}(MN^2 + N^P)$ and $\mathcal{O}(\frac{M}{P}N^{2P})$ respectively.

### 3.4. Full posterior factorization over the parameters

An even faster integration can be obtained by factorizing the posterior as a product of parameter-wise univariate distributions and replacing these distributions by simple parametric forms that are easily integrated analytically or by tabulation. The posterior distributions of log-amplitude factors given other parameters are Gaussian. Equation (17) shows that the posterior distributions of phase parameters are proportional to $\exp(-\|z_{pm}\|^2 \tilde{a}_{pm}\hat{a}_{pm}\sin^2((\phi_{pm}-\tilde{\phi}_{pm})/2)/{\sigma^e}^2)$, thus depending on the single parameter $\|z_{pm}\|^2\tilde{a}_{pm}\hat{a}_{pm}/{\sigma^e}^2$ up to phase rotation. Further analytical computation shows that the posterior distributions of the log-amplitudes of the partials have a complex (possibly multimodal) shape depending on four hyper-parameters: $\tilde{a}_{pm}$, $\hat{r}_p\mu^a_{pm}$, $\sigma^a_p$ and $\|z_{pm}\|/\sigma^e$. Similarly, the posterior distributions of log-fundamental frequencies depend on many hyper-parameters. Therefore the two latter distributions are approximated by simple Gaussians. Integration of these distributions gives

$$P(S|x) \approx P(S,\hat{f},\hat{r},\hat{a},\hat{\phi}|x)\prod_p \mathcal{I}(\hat{c}_{\log f_p})\mathcal{I}(\hat{c}_{\log r_p}) \times \prod_m \mathcal{I}(\hat{c}_{\log a_{pm}})\mathcal{J}(\hat{c}_{\phi_{pm}}), \quad (23)$$

where $\hat{c}_{\log f_p}$, $\hat{c}_{\log r_p}$, $\hat{c}_{\log a_{pm}}$ and $\hat{c}_{\phi_{pm}}$ denote the curvature of the posterior at its maximum with respect to the parameters, defined by $\hat{c}_y = -\partial^2 \log P(S,f,a,r,\phi|x)/\partial y^2$, and

$$\mathcal{I}(c) = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}cy^2}\,dy = (2\pi/c)^{\frac{1}{2}}, \quad (24)$$

$$\mathcal{J}(c) = \int_{-\pi}^{+\pi} e^{-2c\sin^2\frac{y}{2}}\,dy. \quad (25)$$

This equation, resulting in a complexity of $\mathcal{O}(M+P)$, is identical to the diagonal Laplace approximation, except that phase parameters are modeled by their true posteriors instead of Gaussian approximations. Figure 3 suggests that the proposed factorization is generally not exact, but may be more accurate than the diagonal Laplace approximation at phase values far from the optimum.

## 4. EVALUATION

The proposed marginalization algorithms were compared for the task of score transcription on short signal frames without assuming knowledge of the true number of notes. Model hyper-parameters $\sigma^f$, $\mu^a_{pm}$, $\sigma^a_p$, $\mu^r_p$ and $\sigma^r_p$ were learnt on a subset of the RWC Musical Instrument Database[2], while test signals were obtained by selecting and mixing isolated note signals played by five different wind instruments from the University of Iowa Musical Instrument Samples[3]. More precisely, the test set included 100 one-note signals spanning all discrete pitches from $p = 40$ to 87 and 100 two-note signals corresponding to all possible pitch intervals between 1 and 25 semitones with four different lowest pitches $p = 40$, 47, 54 and 61. All signals were sampled at 22.05 kHz and framed with a Hanning window of length 1024 (46 ms). Numerical integration involved $N = 50$ grid points per variable and partials were grouped into subsets according to a difference threshold of 41 Hz, resulting in at most 2 partials per subset.

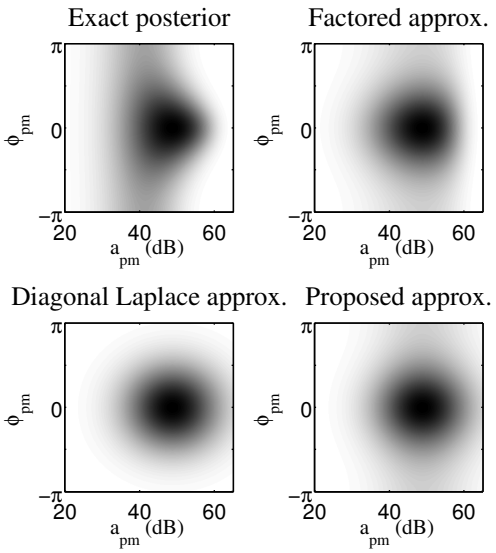



**Fig. 3**. Shape of the joint posterior for a signal containing a partial with 55 dB amplitude and zero phase as a function of the hypothesized amplitude and phase. Dark areas denote high probability. The mean prior amplitude equals 45 dB.

In order to avoid testing all possible states, 6 candidate states (3 with one note and 3 with two notes) were preselected for each test signal as those minimizing the residual of the orthogonal projection of the observed magnitude spectrum on the subspace spanned by the typical magnitude spectra of the active notes, derived from their mean spectral envelopes $\mu^a_{pm}$. The performance was measured in terms of recall $R = N_{\rm cor}/N_{\rm ref}$ and precision $P = N_{\rm cor}/N_{\rm est}$, where $N_{\rm ref}$ is the true number of notes, $N_{\rm est}$ the estimated number of notes and $N_{\rm cor}$ the number of correctly transcribed notes.

The average computation time including optimization was 31 s per candidate for the factorization over subsets of partials, 1.7 s for the factorization over partials, 1.2 s for the full Laplace approximation and 1.1 s for the factorization over parameters and the diagonal Laplace approximation.

Results are shown in table 1. The full Laplace approximation performed worst both for one-note and two-note signals, since unbounded integration of the posterior tended to over-estimate the posterior state probability and select can-

[2]http://staff.aist.go.jp/m.goto/RWC-MDB/

[3]http://theremin.music.uiowa.edu/MIS.html

| Method | 1 note $R/P$ (%) | 2 notes $R/P$ (%) |
|---|---|---|
| Fact. subsets of partials | 100 / 100 | 94.5 / 99.5 |
| Fact. partials | 100 / 100 | 88.5 / 97.8 |
| Fact. parameters | 100 / 100 | 88.0 / 97.8 |
| Diagonal Laplace | 100 / 100 | 88.0 / 98.9 |
| Full Laplace | 94.0 / 54.7 | 74.0 / 80.9 |

**Table 1**. Score transcription performance of the proposed marginalization methods, corresponding to equations (22), (21) and (23) respectively, compared with Laplace methods in terms of recall $R$ and precision $P$.

| Method | Fsub | Fprt | Fprm | DLap | FLap |
|---|---|---|---|---|---|
| Fsub | N/S | | | | |
| Fprt | $5\times10^{-4}$ | N/S | | | |
| Fprm | $2\times10^{-4}$ | 1 | N/S | | |
| DLap | $2\times10^{-4}$ | 1 | N/S | N/S | |
| FLap | $1\times10^{-10}$ | $3\times10^{-5}$ | $4\times10^{-5}$ | $4\times10^{-5}$ | N/S |

**Table 2**. McNemar statistics of the marginalization methods (sorted as in table 1) on two-note signals. Values smaller than 0.05 indicate significantly different performance, while N/S indicates empirically identical performance.

didates with a larger number of partials. All other algorithms provided perfect transcription on one-note signals. The factorization over subsets of partials lead to the best performance on two-note signals, while other factorizations and the diagonal Laplace approximation achieved similarly lower performances. McNemar statistics [11] displayed in table 2 support the significance of these conclusions. This suggests that dependencies between the parameters of different partials must be taken into account to achieve a good performance on multi-note signals, while the distribution of the parameters of each partial can be approximated by simpler factored distributions without consequence.

## 5. CONCLUSION

We investigated several factorizations of the posterior distribution of the parameters of a harmonic model and exploited them for Bayesian inference. Analytical computation resulted in an exact conditional factorization over subsets of partials with close frequencies. Further approximations leading to tractable inference were proposed by removing some conditional dependencies. These factorizations rely on the fact that the dependencies between the partials are modeled using a limited number of parameters, namely a fundamental frequency and a global amplitude parameter. Thus they could be applied to other harmonic models following this assumption. Score transcription experiments showed that all factorizations performed perfectly on one-note signals but that the non-modeling of dependencies between parameters of different partials degraded the performance on two-note signals.

In the future, we plan to improve the computational efficiency of inference based on the factorization of the posterior over subsets of partials. Replacing numerical integration over each subset of parameters by MCMC sampling when appropriate seems a promising approach. We think this could potentially result in a faster inference than straightforward MCMC with only a small performance decrease.

## 6. REFERENCES


[1] D. P. W. Ellis, *Prediction-driven computational auditory scene analysis*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1996.

[2] M. P. Ryynänen and A. P. Klapuri, "Polyphonic music transcription using note event modeling," in *Proc. WASPAA*, 2005, pp. 319–322.

[3] C. Raphael, "Automatic transcription of piano music," in *Proc. ISMIR*, 2002, pp. 15–19.

[4] E. Vincent, "Musical source separation using time-frequency source priors," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 91–98, 2006.

[5] P. J. Walmsley, S. J. Godsill, and P. J. W. Rayner, "Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters," in *Proc. WASPAA*, 1999, pp. 119–122.

[6] M. Davy and S. J. Godsill, "Bayesian harmonic models for musical pitch estimation and analysis," Tech. Rep. CUED/F-INFENG/TR.431, Cambridge University, Dept. of Engineering, Cambridge, UK, 2002.

[7] S. J. Godsill and M. Davy, "Bayesian computational models for inharmonicity in musical instruments," in *Proc. WASPAA*, 2005, pp. 283–286.

[8] G. Casella and C. P. Robert, *Monte Carlo Statistical Methods, 2nd Edition*, Springer, 2005.

[9] E. Vincent and M. D. Plumbley, "A prototype system for object coding of musical audio," in *Proc. WASPAA*, 2005, pp. 239–242.

[10] D. M. Chickering and D. Heckerman, "Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables," in *Proc. UAI*, 1996, pp. 158–168.

[11] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures, 2nd Edition*, Chapman & Hall, 2000.